

1). Which are the top three variables in your model that contribute most towards the probability of a lead getting converted?

Lead Source: The origin of the lead (e.g., email campaign, website visit).

Lead Activity: Actions like page views or form submissions.

Lead Quality Score: A manually or automatically assigned score.

These can be identified through Logistic Regression coefficients, feature importance in Decision Trees, or Recursive Feature Elimination (RFE).

1. 2). What are the top 3 categorical/dummy variables in the model which should be focused the most on in order to increase the probability of lead conversion?

Lead Source: Certain sources, like email campaigns or direct website inquiries.

Lead Stage: Leads further along in the funnel (e.g., "Interested" or "Qualified").

Industry/Job Role: Leads from specific industries or roles with a higher propensity to convert.

3. X Education has a period of 2 months every year during which they hire some interns. The sales team, in particular, has around 10 interns allotted to them. So during this phase, they wish to make the lead conversion more aggressive. So they want almost all of the potential leads (i.e. the customers who have been predicted as 1 by the model) to be converted and hence, want to make phone calls to as much of such people as possible. Suggest a good strategy they should employ at this stage.

```
build_model_cutoff(X_train[col], y_train, X_test[col], y_test, cutoff=0.1)
```

-----Result of training data-----

	Prospect ID	Converted	Convert_Probability	Convert_predicted	Lead_Score
0	3009	0	0.297269	1	30
1	1012	0	0.017253	0	2
2	9226	0	0.318173	1	32
3	4750	1	0.683229	1	68
4	7987	1	0.761301	1	76

-----Result of test data-----

	Prospect ID	Converted	Convert_Probability	Convert_predicted	Lead_Score
0	3271	0	1.000000	1	100
1	1490	1	1.000000	1	100
2	7936	0	1.000000	1	100
3	4216	1	0.861157	1	86
4	3830	0	1.000000	1	100

-----Model Evaluation Metrics-----

Confusion Matrix :

```
[[ 11 1723]
 [ 2 987]]
```

Accuracy : 0.36650752846125595

Sensitivity : 0.9979777553083923

Specificity : 0.006343713956170703

Precision : 0.36420664206642067

	Prospect ID	Converted	Convert_Probability	Convert_predicted	Lead_Score
0	3271	0	1.000000	1	100
1	1490	1	1.000000	1	100
2	7936	0	1.000000	1	100
3	4216	1	0.861157	1	86
4	3830	0	1.000000	1	100
...
2718	850	0	1.000000	1	100
2719	2879	0	1.000000	1	100
2720	6501	1	1.000000	1	100
2721	7155	0	1.000000	1	100
2722	376	0	1.000000	1	100

2723 rows × 5 columns

- Similarly, at times, the company reaches its target for a quarter before the deadline. During this time, the company wants the sales team to focus on some new work as well. So during this time, the company's aim is to not make phone calls unless it's extremely necessary, i.e. they want to minimize the rate of useless phone calls. Suggest a strategy they should employ at this stage.

```
build_model_cutoff(X_train[col], y_train, X_test[col], y_test, cutoff=0.9)
```

-----Result of training data-----

	Prospect ID	Converted	Convert_Probability	Convert_predicted	Lead_Score
0	3009	0	0.297269	0	30
1	1012	0	0.017253	0	2
2	9226	0	0.318173	0	32
3	4750	1	0.683229	0	68
4	7987	1	0.761301	0	76

-----Result of test data-----

	Prospect ID	Converted	Convert_Probability	Convert_predicted	Lead_Score
0	3271	0	1.000000	1	100
1	1490	1	1.000000	1	100
2	7936	0	1.000000	1	100
3	4216	1	0.861157	0	86
4	3830	0	1.000000	1	100

-----Model Evaluation Metrics-----

Confusion Matrix :

```
[[ 415 1319]
```

```
 [ 243  746]]
```

Accuracy : 0.4263679764965112

Sensitivity : 0.7542972699696663

Specificity : 0.23933102652825836

Precision : 0.3612590799031477

	Prospect ID	Converted	Convert_Probability	Convert_predicted	Lead_Score
0	3271	0	1.000000	1	100
1	1490	1	1.000000	1	100
2	7936	0	1.000000	1	100
3	4216	1	0.861157	0	86
4	3830	0	1.000000	1	100
...
2718	850	0	1.000000	1	100
2719	2879	0	1.000000	1	100
2720	6501	1	1.000000	1	100
2721	7155	0	1.000000	1	100
2722	376	0	1.000000	1	100

2723 rows × 5 columns